NeuroImage 244 (2021) 118618

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Empirical Bayes estimation of pairwise maximum entropy model for nonlinear brain state dynamics

Seok-Oh Jeong [a], Jiyoung Kang [b,c], Chongwon Pae [b,c], Jinseok Eo [b,c,d], Sung Min Park [b], Junho Son [b,c,d], Hae-Jeong Park [b,c,d,e,*]

[a] Department of Statistics, Hankuk University of Foreign Studies, Yong-In, Republic of Korea
[b] Center for Systems and Translational Brain Science, Institute of Human Complexity and Systems Science, Yonsei University, Seoul, Republic of Korea
[c] Department of Nuclear Medicine, Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea
[d] Graduate School of Medical Science, Yonsei University College of Medicine, Brain Korea 21 Project, Seoul, Republic of Korea
[e] Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Maximum entropy model
Resting state
Brain dynamics
Energy landscape
Variational Bayes
Variational expectation-maximization
Empirical Bayes
Hierarchical Bayesian parameter estimation

ABSTRACT

The pairwise maximum entropy model (pMEM) has recently gained widespread attention to exploring the nonlinear characteristics of brain state dynamics observed in resting-state functional magnetic resonance imaging (rsfMRI). Despite its unique advantageous features, the practical application of pMEM for individuals is limited as it requires a much larger sample than conventional rsfMRI scans. Thus, this study proposes an empirical Bayes estimation of individual pMEM using the variational expectation-maximization algorithm (VEM-MEM). The performance of the VEM-MEM is evaluated for several simulation setups with various sample sizes and network sizes. Unlike conventional maximum likelihood estimation procedures, the VEM-MEM can reliably estimate the individual model parameters, even with small samples, by effectively incorporating the group information as the prior. As a test case, the individual rsfMRI of children with attention deficit hyperactivity disorder (ADHD) is analyzed compared to that of typically developed children using the default mode network, executive control network, and salient network, obtained from the Healthy Brain Network database. We found that the nonlinear dynamic properties uniquely established on the pMEM differ for each group. Furthermore, pMEM parameters are more sensitive to group differences and are better associated with the behavior scores of ADHD compared to the Pearson correlation-based functional connectivity. The simulation and experimental results suggest that the proposed method can reliably estimate the individual pMEM and characterize the dynamic properties of individuals by utilizing empirical information of the group brain state dynamics.

1. Introduction

The human brain is considered as a nonlinear dynamic system that transitions multiple stable states (Breakspear, 2017; Cabral et al., 2014; Deco and Jirsa, 2012; Deco et al., 2015; Freyer et al., 2011, 2012; Kelso, 2012; Rabinovich and Varona, 2011; Tognoli and Kelso, 2014). The nonlinear dynamics of the brain states transitioning among the stable states (local minima, called "attractors") have been explored in terms of the energy landscape of brain states. The energy of a state is defined as the negative log probability of the occurrence of the state. The energy landscape contains various states, such as stable states, transition states, and saddle states. From a network perspective of the brain, these states occur due to the nonlinear interactions among distributed regions of the brain and not as a result of random. A pairwise maximum entropy model (pMEM) is generally used to link the nonlinear interactions and the occurrence of each state (i.e., energy). Until now, the energy landscape analysis based on the pMEM has been studied mainly for the brain during the resting state (Ezaki et al., 2020; Kang et al., 2017, 2019; Watanabe et al., 2013, 2014a). Energy landscape analysis has identified the abnormal state dynamics in the brain diseases such as autism spectrum disorder (Watanabe and Rees, 2017), observed in the resting-state functional magnetic resonance imaging (rsfMRI). Park and Kang (2021) used the nonlinear dynamic states to represent the long-term interaction with the environment. In these studies, the brain states have been defined with binarized activity in the distributed brain regions (Ezaki et al., 2018; Gu et al., 2018; Kang et al., 2017; Watanabe et al., 2013, 2014b, 2014c, 2014d).

Despite all the advantages of pMEM for modeling nonlinear interaction of the brain, it requires a large sample to estimate the model pa-

* Corresponding author at: Graduate School of Medical Science, Department of Nuclear Medicine, Yonsei University, College of Medicine, 50-1 Yonsei-ro, Sinchon-dong, Seodaemun-gu, Seoul 03722, Republic of Korea.
E-mail address: parkhj@yonsei.ac.kr (H.-J. Park).

https://doi.org/10.1016/j.neuroimage.2021.118618.
Received 13 May 2021; Received in revised form 22 September 2021; Accepted 24 September 2021
Available online 25 September 2021.
1053-8119/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

rameters reliably. It is not easy to obtain a long time series from an individual for the rsfMRI, particularly in the clinical settings (approximately 300 sample points or 5–10 min per person). Furthermore, most of the estimation procedures for the pMEM, such as the maximum likelihood estimation (MLE) method, assume that the sample points are independently observed. Since the rsfMRI samples are temporally correlated due to the prolonged hemodynamic response functions (Kang et al., 2021), the original sample is subsampled to avoid the temporal dependence, so that the time series gets shorter. Therefore, the pMEM analysis with rsfMRI has been primarily conducted at the group level by concatenating the time series of the group data. The individual variation cannot be considered in this group-concatenation approach. To overcome the small sample size in characterizing the dynamic properties of each individual, Kang et al. (2021) developed a Bayesian model parameter estimation scheme for the pMEM using a variational Bayes approach, where the group concatenated data are used to determine the prior mean. However, the prior precision should be determined manually due to the absence of an automatic device for selecting the prior precision empirically.

This study proposes a new empirical Bayes approach to estimate all the hyperparameters for the prior as well as the individual pMEM model parameters in a fully automatic and data-driven manner by utilizing the information from the group data. To empirically construct a prior for the model parameters, we used an empirical Bayes approach by considering the unknown individual pMEM parameters as a realization of a multivariate random variable instead of a set of fixed values. The randomness of these quantities is expressed as the prior distribution, usually representing the uncertainty associated with their values. We regarded the uncertainty as arising from the heterogeneity among subjects in the group under a study. To determine the uncertainty as well as the group pMEM values, we adopted the variational approximation method and developed an expectation-maximization (EM)-like algorithm for the variational approximation, called the variational EM algorithm for pairwise MEM (VEM-MEM). This method learns the prior in a fully data-driven way so that the 'suitableness' of the prior is supported by the prior information provided by subjects' data in the group. We expect a group prior works as a device keeping the parameter estimates from erratic values. The resulting posterior can be eligible for the inference of the individual parameters. Thus, the VEM-MEM algorithm yields a solution by alternating two-level processes, i.e., the individual-level and group-level processes. At the individual level, the individual pMEM parameter estimate is given by correcting the prior (group) mean vector by the deviation of the empirical mean of the observed individual data from its counterpart under the (group) prior distribution. In this level, the individual precision is the sum of the group precision and the information amount added by the observed data from the subject. At the group level, the prior (group) mean vector is given by the average of the individual pMEM parameters. The prior precision at the group level is incorporated with the average of all individual precisions and the heterogeneity of the individual pMEM parameters.

Simulation studies with various setups are conducted to evaluate the performance of the developed methodology in comparison with conventional maximum likelihood estimation (MLE). The proposed method can be practically used for reasonable network size, around 10–20 nodes with clinically available rsfMRI samples, to analyze the energy landscape of brain states.

To test the applicability of the proposed method to a group study based on the individualized estimation of the pMEM and to show the usefulness of the energy landscape features in characterizing individuals or groups, we applied the proposed pMEM estimation and energy landscape analysis to a clinical case. As an exemplary clinical case, we analyzed the differential state dynamics in typically developed children (TDC) and children with attention deficit hyperactivity disorder (ADHD), which is a complex neurobehavioral disorder commonly observed in childhood, using the rsfMRI data from the Healthy Brain Network (HBN) database (Alexander et al., 2017). Not only the technical improvement, this study extends the need of individualized pMEM presented in Kang et al. (2021) to a group study; Kang et al. (2021) showed the usefulness of pMEM and energy landscape features in characterizing cognitive properties. Recent neuroimaging studies on ADHD have suggested the effect of dysfunction in brain connectivity on ADHD (see Saad et al., 2020 for review). Accordingly, brain connectivity has been analyzed in relation to ADHD, mainly focusing on the default mode network (DMN) (Anderson et al., 2014; Barber et al., 2015; Brown et al., 2012; Fair et al., 2010; Qiu et al., 2011; Sonuga-Barke and Castellanos, 2007; Zang et al., 2007), the executive control network (ECN) (Cao et al., 2006; Hilger and Fiebach, 2019; Konrad et al., 2006; Qiu et al., 2011), and the salience network (SAN) (Cai et al., 2015; Hilger and Fiebach, 2019; Sidlauskaite et al., 2016) corresponding to the cognitive deficits in ADHD. We used these subnetworks to evaluate the VEM-MEM in characterizing the differential state dynamics in ADHD, reported in the dynamic state transition analysis (Scofield et al., 2019; Shappell et al., 2021).

This study is divided into three parts. Firstly, we present a mathematical basis for the VEM-MEM. Secondly, we evaluate the VEM-MEM scheme using simulation studies. Thirdly, we apply the VEM-MEM to the rsfMRI of the ADHD dataset and evaluate the applicability of the pMEM in exploring mental disorders. By simulation and application to the ADHD dataset, we showed that VEM-MEM could reliably estimate the pMEM parameters with the conventionally obtained rsfMRI samples.

2. Methods

2.1. The pairwise maximum entropy model

Let us begin with a brief introduction of pMEM, which is an extension of the formulation in Kang et al. (2021). For a detailed mathematical review of pMEM, refer to Yeh et al. (2010).

Suppose the brain state space S is represented by

$$S = \left\{\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_d)^\top \in \{0,1\}^d | \boldsymbol{\sigma} \text{ is a possible state}\right\}$$

where the value of σ_i $(i = 1, 2, ..., d)$ is either 0 (inactive) or 1 (active), indicating a local activity at a node (brain region) i, and d is the total number of nodes (or ROIs). It is well known that maximizing the entropy of the state space results in a Boltzmann distribution having a probability mass function

$$p(\boldsymbol{\sigma}|\boldsymbol{h},\boldsymbol{J}) = \frac{\exp\{-E(\boldsymbol{\sigma}|\boldsymbol{h},\boldsymbol{J})\}}{\sum_{\boldsymbol{\sigma}'\in S}\exp\left\{-E\left(\boldsymbol{\sigma}'|\boldsymbol{h},\boldsymbol{J}\right)\right\}}$$

where

$$E(\boldsymbol{\sigma}|\boldsymbol{h},\boldsymbol{J}) = -\sum_{i=1}^{d} h_i\sigma_i - \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} J_{ij}\sigma_i\sigma_j \text{ with } \boldsymbol{h} = \left(h_i\right)_{1:d} \text{ and } \boldsymbol{J} = \left(J_{ij}\right)_{d\times d}$$

is the energy and h_i's and J_{ij}'s are the model parameters, indicating the excitability (or sensitivity) at the node i and the interaction between the nodes i and j. The total number of parameters in this model is then $D = d + d(d-1)/2 = d(d+1)/2$. For convenience, we reparametrize the model $p(\boldsymbol{\sigma}|\boldsymbol{h},\boldsymbol{J})$ as follows:

$$p(\boldsymbol{\sigma}|\boldsymbol{\theta}) = \exp\left\{\boldsymbol{\theta}^\top\tilde{\boldsymbol{\sigma}} - \log\sum_{\boldsymbol{\sigma}'\in S}\exp\left(\boldsymbol{\theta}^\top\tilde{\boldsymbol{\sigma}}'\right)\right\}$$

where $\tilde{\boldsymbol{\sigma}} = (\sigma_1, \cdots, \sigma_d, \sigma_1\sigma_2, \cdots, \sigma_{d-1}\sigma_d)^\top \in \{0,1\}^D$, $\boldsymbol{\theta} = (h_1, \cdots, h_d, J_{12}, \cdots, J_{d-1,d})^\top \in \mathbb{R}^D$ and the symbol $\top$ denotes the matrix transpose.

2.2. Bayesian formulation and variational EM algorithm

Suppose we have datasets

$$\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_N$$

observed from N independent subjects. Let $\mathcal{D}$ be the pooled dataset:

$$\mathcal{D} = \cup_{n=1}^{N}\mathcal{D}_n$$

Then, the log-likelihood function of the pooled dataset D is given by simply summing up the log-likelihoods $\ell(\theta_n) = \log p(D_n|\theta_n)$ over all subjects:

$$\ell(\theta_1, \dots, \theta_N) \equiv \log p(D|\theta_1, \dots, \theta_N) = \sum_{n=1}^{N} \log p(D_n|\theta_n) = \sum_{n=1}^{N} \left[\sum_{t=1}^{T_n} \theta_n^\top \tilde{\sigma}_n(t) - T_n \log \sum_{\sigma \in S} e^{\theta_n^\top \tilde{\sigma}} \right]$$

where θ_n is the model parameter for the individual subject $n = 1, 2, ..., N$. Note that the resulting maximum likelihood estimates for θ_n's are the same as those composed of individual MLE's obtained by maximizing $\ell(\theta_n)$ separately for each $n = 1, 2, ..., N$, which means that, under MLE scheme, there's no way for the estimation procedure for a specific subject to learn from the other subjects.

We consider the subject-specific model parameter vector $\theta_n \in \mathrm{R}^D (n = 1, 2, ..., N)$ as a realization from the prior distribution $N(\eta, \mathrm{diag}(\alpha)^{-1})$ with $\eta = (\eta_1, ..., \eta_D)^\top \in \mathrm{R}^D$ and $\alpha = (\alpha_1, ..., \alpha_D)^\top \in \mathrm{R}_+^D$. Put

$$\Theta = (\theta_{nj})_{N \times D} = \left[\theta_1, \theta_2, \cdots, \theta_N\right]^\top,$$

so that θ_{nj} denotes the j-th component of the model parameter vector θ_n $(n = 1, 2, ..., N;\ j = 1, 2, \dots, D)$. Then the prior for Θ is given by $p(\Theta|\eta, \alpha) = \prod_{n=1}^{N} N(\eta, \mathrm{diag}(\alpha)^{-1})$, assuming that the subjects are mutually independent.

Now we are to obtain the posterior $p(\Theta|D, \eta, \alpha)$ combining the normal prior $p(\Theta|\eta, \alpha)$ with the data likelihood $p(D|\Theta) = \prod_{n=1}^{N} p(D_n|\theta_n)$:

$$p(\Theta|D, \eta, \alpha) = \frac{p(\Theta|\eta, \alpha) p(D|\Theta)}{p(D)}$$

where $p(D) = \int p(D|\Theta) p(\Theta|\eta, \alpha) d\Theta$ is the evidence provided by the observed data.

However, it is intractable to derive the posterior analytically, since the normal distribution is not a conjugate prior for the Boltzmann distribution and it is impossible to obtain the closed form of $p(D)$. So, we consider a variational approximation for the posterior by the multivariate normal distribution as follows:

$$p(\Theta|D, \eta, \alpha) \approx q(\Theta) = \prod_{n=1}^{N} N\left(\mu_n, \mathrm{diag}\left(\beta_n\right)^{-1}\right)$$

where $\mu_n = (\mu_{n1}, ..., \mu_{nD})^\top \in \mathrm{R}^D$ and $\beta_n = (\beta_{n1}, ..., \beta_{nD})^\top \in \mathrm{R}_+^D$ are the posterior mean vector and the posterior precision vector for the subject $n = 1, 2, .., N$, respectively. Then, the resulting μ_n will be the MAP estimate for the model parameter θ_n for $n = 1, 2, .., N$.

The variational approximation $q(\Theta)$ is given by the minimizer of Kullback-Leibler divergence between $q(\Theta)$ and $p(\Theta|D, \eta, \alpha)$

$$\mathrm{KL}(q\|p) = \int q(\Theta) \log \frac{p(\Theta|D, \eta, \alpha)}{q(\Theta)} d\Theta$$

among a class of candidate distributions (in this study, the class of normal distributions). Note that the log-evidence, $\log p(D)$, is decomposed as follows:

$$\log p(D) = KL(q||p) + \mathcal{F}(q)$$

where $\mathcal{F}(q) \equiv \mathcal{F}(q|\eta, \alpha) = \int q(\Theta) \log \frac{p(D, \Theta|\eta, \alpha)}{q(\Theta)} d\Theta$ is the free-energy. Since the evidence $p(D)$ is not related with any variational parameters, the variational solution which minimizes $KL(q||p)$ coincides with the maximizer of $\mathcal{F}(q)$. Note that, since the Kullback-Leibler divergence is always nonnegative, the free-energy constitutes the evidence lower bound (ELBO).

In this study, we propose a variational EM procedure to obtain the values of hyperparameters (η, α) empirically as well as the posterior parameters (μ_n, β_n)'s. For this, consider the decomposition of ELBO as follows:

$$\mathcal{F}(q|\eta, \alpha) = \int q(\Theta) \log \frac{p(D, \Theta|\eta, \alpha)}{q(\Theta)} d\Theta = \int q(\Theta) \log \frac{p(D|\Theta) p(\Theta|\eta, \alpha)}{q(\Theta)} d\Theta = \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3$$

with $\mathcal{F}_1 = \int q(\Theta) \log p(D|\Theta) d\Theta$, $\mathcal{F}_2 = \int q(\Theta) \log p(\Theta|\eta, \alpha) d\Theta$, and $\mathcal{F}_3 = -\int q(\Theta) \log q(\Theta) d\Theta$. Since the integration in $\mathcal{F}_1$ is intractable, we consider a second-order Taylor series expansion near $\theta_n = \eta$ for the integrand using the similar technics as in **Appendix A1** of Kang et al. (2021), so that we have the approximation $\mathcal{F}_1 \approx \tilde{\mathcal{F}}_1$ as follows:

$$\begin{aligned} \tilde{\mathcal{F}}_1 = & \sum_{n=1}^{N} \sum_{t=1}^{T_n} \mu_n^\top \tilde{\sigma}_n(t) - \sum_{n=1}^{N} T_n \log \sum_{\sigma \in S} e^{\eta^\top \tilde{\sigma}} - \sum_{n=1}^{N} T_n \cdot \langle \tilde{\sigma} \rangle_\eta^\top (\mu_n - \eta) \\ & - \frac{1}{2} \sum_{n=1}^{N} T_n \left\{ \mathrm{tr}\left(\mathrm{diag}\left(\beta_n\right)^{-1} C_\eta \right) + (\mu_n - \eta)^\top C_\eta (\mu_n - \eta) \right\} \end{aligned}$$

with $C_\eta = Cov_\eta(\tilde{\sigma}) = \langle \tilde{\sigma}\, \tilde{\sigma}^\top \rangle_\eta - \langle \tilde{\sigma} \rangle_\eta \langle \tilde{\sigma} \rangle_\eta^\top$. $\mathcal{F}_2$ and $\mathcal{F}_3$ are simply given by

$$\mathcal{F}_2 = \frac{N}{2} \mathbf{1}^\top (\log \alpha) - \frac{1}{2} \sum_{n=1}^{N} \left\{ (\mu_n - \eta)^\top \mathrm{diag}(\alpha) (\mu_n - \eta) + \alpha^\top (1/\beta_n) \right\} + \text{constant}$$

and

$$\mathcal{F}_3 = -\frac{1}{2} \sum_{n=1}^{N} \mathbf{1}^\top \log \beta_n + \text{constant}.$$

Now we are ready to state our variational EM procedure precisely. For the E-step, we are to maximize

$$\tilde{\mathcal{F}}(q|\eta, \alpha) = \tilde{\mathcal{F}}_1 + \mathcal{F}_2 + \mathcal{F}_3$$

with respect to q holding η and α fixed. This step gives, for $n = 1, 2, \dots, N$,

$$\mu_n = \eta + \left\{ T_n C_\eta + \mathrm{diag}(\alpha) \right\}^{-1} T_n \left(\langle \tilde{\sigma}_n \rangle_{\mathrm{emp}} - \langle \tilde{\sigma} \rangle_\eta \right)$$

and

$$\beta_n = T_n c_\eta + \alpha,$$

where c_η is the vector composed of the diagonal elements of C_η. Next, for the M-step, holding q fixed, we are to maximize $\mathcal{F}$(not $\tilde{\mathcal{F}}$) with respect to η and α. Then we have

$$\eta = \frac{1}{N} \sum_{n=1}^{N} \mu_n$$

and

$$1/\alpha = \frac{1}{N} \sum_{n=1}^{N} \left\{ (\mu_n - \eta) \odot (\mu_n - \eta) + 1/\beta_n \right\},$$

where $\odot$ denotes the Hadamard product operator.

Hence, the variational EM algorithm for MEM is summarized as follows:

1 (Initialize) Set $\eta = \eta_0$, $\alpha = \alpha_0$.
2 (E-step: individual-level, update q)

$$\mu_n = \eta + \left\{ T_n C_\eta + \mathrm{diag}(\alpha) \right\}^{-1} T_n \left(\langle \tilde{\sigma}_n \rangle_{\mathrm{emp}} - \langle \tilde{\sigma} \rangle_\eta \right)$$

and

$$\beta_n = T_n c_\eta + \alpha$$

for $n = 1, 2, ..., N$
3 (M-step: group-level, update η and α)

$$\eta = \frac{1}{N} \sum_{n=1}^{N} \mu_n$$

and

$$\alpha = \left[\frac{1}{N} \sum_{n=1}^{N} \left\{ (\mu_n - \eta) \odot (\mu_n - \eta) + 1/\beta_n \right\} \right]^{-1}.$$

4 If converged, then STOP. Otherwise, go to the step 2.

Remark It is noticeable that, unlike the maximum likelihood scheme, the proposed variational EM algorithm enables the estimation procedure for a specific subject model parameter to learn from other datasets. While the maximum likelihood method uses only T_n time points to estimate θ_n for each subject $n = 1, 2, ..., N$ separately, the variation EM

procedure allows all $T = \sum_{n=1}^{N} T_n$ observations contribute to estimate the individual model parameter θ_n for $n = 1, 2, ..., N$ through η and α.

Remark Practically, it would be rather helpful to assume that the prior precision α is represented by two scalars $\alpha_h > 0$ and $\alpha_J > 0$ only:

$$\boldsymbol{\alpha} = \left(\alpha_h \mathbf{1}_d^\top, \alpha_J \mathbf{1}_{D-d}^\top\right)^\top$$

where α_h is the precision for $\boldsymbol{h}$, α_J is the precision for $\boldsymbol{J}$, and $\mathbf{1}_K$ is the vector of length K with all components equal to one. Under this restricted setting, it is straightforward to verify the update rule for the prior precision $\boldsymbol{\alpha}$ in the M-step is modified as follows:

$$\alpha_h = \left[\frac{1}{N}\sum_{n=1}^{N} \overline{\left\{\left(\boldsymbol{\mu}_{n,1:d} - \boldsymbol{\eta}_{1:d}\right) \odot \left(\boldsymbol{\mu}_{n,1:d} - \boldsymbol{\eta}_{1:d}\right) + 1/\boldsymbol{\beta}_{n,1:d}\right\}}\right]^{-1},$$

$$\alpha_J = \left[\frac{1}{N}\sum_{n=1}^{N} \overline{\left\{\left(\boldsymbol{\mu}_{n,(d+1):D} - \boldsymbol{\eta}_{(d+1):D}\right) \odot \left(\boldsymbol{\mu}_{n,(d+1):D} - \boldsymbol{\eta}_{(d+1):D}\right) + 1/\boldsymbol{\beta}_{n,(d+1):D}\right\}}\right]^{-1}$$

where the overline denotes the average of all vector components.

2.3. Simulation studies

To check whether the proposed variational EM algorithm works as expected, we conducted simulation studies with the setting $N = 20$ and $d = 10$ under a scenario as follows.

1. (Hyperparameters for the prior) The prior mean vector η was made of $D = d(d+1)/2$ pseudo random numbers generated from the normal distribution having the mean 0 and the standard deviation 0.1. And the prior precision vector was set to be 5 for *h's* and 25 for *J's*, that is, $\boldsymbol{\alpha} = (5 \cdot \mathbf{1}_d^\top, 25 \cdot \mathbf{1}_{D-d}^\top)^\top$.
2. (Model parameters and dataset for each subject) For each $n = 1, 2, ..., N$, θ_n was generated from the multivariate normal distribution $N(\boldsymbol{\eta}, \mathrm{diag}(\boldsymbol{\alpha}))^{-1})$, and then D_n was generated from the corresponding Boltzmann distribution with the model parameter θ_n.
3. (Parameter estimation) Using the simulated datasets, we applied the maximum likelihood method and the variational EM algorithm to estimate the model parameter vector for each subject. Then, the estimates were compared with the true model parameters generated in the step 2.
4. (Evaluation) The estimation error for each subject is measured by the Euclidean distance between the estimated parameter vector $\hat{\theta}_n$ and the true model parameter θ_n defined as follows: $D(\hat{\theta}_n, \theta_n) = \sqrt{(\hat{\theta}_n - \theta_n)^\top (\hat{\theta}_n - \theta_n)}$ for each subject $n = 1, 2, ..., N$. And then, average the squared distances over N subjects to get the overall discrepancy $\overline{D^2} = \frac{1}{N}\sum_{n=1}^{N} D(\hat{\theta}_n, \theta_n)^2 = \frac{1}{N}\sum_{n=1}^{N} (\hat{\theta}_n - \theta_n)^\top (\hat{\theta}_n - \theta_n)$.
5. (Comparison) Compare the overall discrepancy of MLE with that of VEM-MEM using the performance measures in the step 4.

This scenario was implemented for the settings with sample sizes $T_n = 25$, $T_n = 50$ and $T_n = 100$.

2.4. Monte Carlo simulation

To check whether the findings from the simulation study in the previous subsection can be generalized, we performed Monte Carlo experiments with 100 replication with various setups ($d = 5,\ 10$; $N = 20,\ 40,\ 60$; $T_n = 25,\ 50,\ 100$) and the empirical distributions of overall discrepancies $\overline{D^2}$ were investigated.

2.5. Experimental study

The dataset used in this study was obtained from Release 2 and 3 of the Child Mind Institute's HBN Biobank (http://fcon_1000.projects.nitrc.org/). The subjects were clinically assessed and underwent MR scanning sessions based on the HBN's Serial Scanning Initiative protocol (see https://fcon_1000.projects.nitrc.org/indi/hbn_ssi/mri_protocol.html). We selected data of the control children and the children who were diagnosed with either ADHD-Inattentive or ADHD-Combined types based on the clinical assessment. The rsfMRI and the T1-weighted structural MRI were screened, and the data with excessive motion or poor data quality were excluded. The resulting dataset included a total of 298 children (205 males and 93 females) with ages ranging from 5 to 20. This included 233 ADHD children (mean age, 10.9 years; standard deviation, 3.05 years; 172 males and 61 females) and 65 control children (mean age, 10.37 years; standard deviation, 3.19 years; 33 males and 32 females). Among diverse subtypes of ADHD, we only included the two most prominent subtypes: 134 ADHD-Inattentive children and 99 ADHD-Combined type children. A detailed description of the MRI parameters can be found in Alexander et al. (2017).

A standard preprocessing pipeline was conducted for the fMRI data using SPM12 (http://www.fil.ion.ucl.ac.uk/spm, Wellcome Trust Centre for Neuroimaging, London, UK) (Friston et al., 1995) to analyze the pMEM. All the fMRI data were subjected to slice scan time correction, head motion correction, co-registration of T1-weighted images to the first EPI, and spatial normalization to convert the rsfMRI into the MNI template space using nonlinear transformation.

After preprocessing the data, ICA was performed using the Group ICA of the fMRI Toolbox (GIFT) (https://trendscenter.org/software/gift/) (Calhoun et al., 2001). The number of independent components was estimated as 41 based on the minimum description length (MDL) criteria (Li et al., 2007). Using the group independent components (ICs), the subject-specific spatial maps and corresponding time courses were predicted using back-reconstruction. The subject time courses were then converted to the z-scores. The components were visually inspected and manually labeled according to the cluster locations: five core components were classified as the DMN, three as SAN components, and three as ECN components. We constructed an ADHD-related network by combining DMN, SAN, and ECN (a total of 11 components), called 'ADHDN'. Since the subnetwork sizes of the SAN and ECN are small, we combined those networks into a subnetwork called 'ECN+SAN' (a total of 6 components).

The component weight time series were zero-thresholded to binarize the component activity states to represent specific brain states. All the (binarized) component states together comprise a state vector. For the analysis of the state dynamics of the three individual subnetworks, the energy landscapes based on the estimated pMEM were constructed following Kang et al. (2017). The distance between two state vectors is defined as the number of elements (bits) that differ from each other. Assuming a gradual state transition, the energy landscape of the brain state is explored by changing each element of the state vector. Based on this process, the local minima (LM), the occupation times (OCC), and the basin sizes of local minima are calculated. The local minima (i.e., stable states) indicate the states with lower energy compared to their neighbors. The basin of a local minimum refers to all the states with the gradients pointed toward the local minimum. The fraction of the states that belong to the basin of a local minimum is the basin size of the local minimum. The occupation time of each local minimum is calculated as the sum of the probabilities of the states in its basin region. Fig. 1 illustrates the procedure described above.

Among the local minima, we focused on those with the first and the second-lowest energies (LM1 and LM2). The number of the local minima, the occupation time of the LM1 (OCC1) and the sum of the occupation times (OCC1+2) of the LM1 and LM2 were chosen as the features of interest in the group comparison. Note that OCC1 and OCC1+2 indicate the occurrence portions of the primary state and the primary and secondary states, respectively. The group difference in these features observed in the three subnetworks (DMN, ECN+SAN, and ADHDN) were investigated by the one-way analysis of covariance (ANCOVA), with the age and the sex being the covariates. We also conducted ANCOVA for

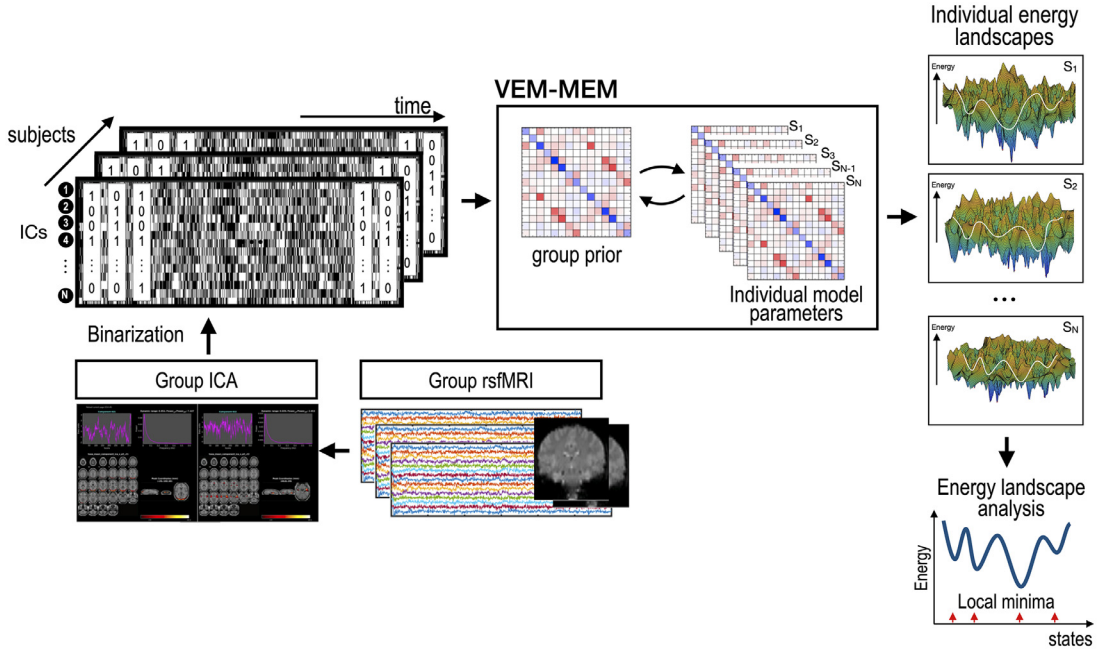

Fig. 1. An illustration of energy landscape analysis procedure.

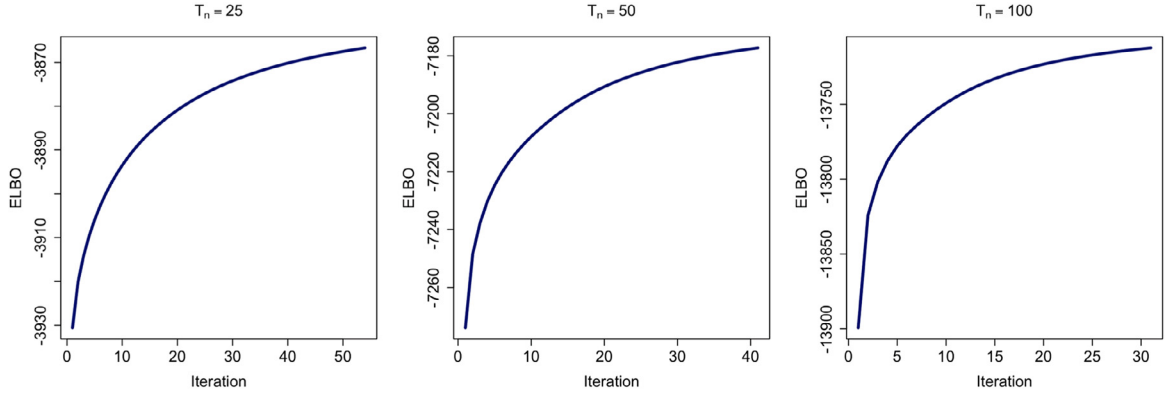

Fig. 2. The trace of ELBO after each variational EM iteration step for simulated datasets ($N = 20$, $d = 10$, $T_n = 25,\ 50,\ 100$). The algorithm converged well in all cases. As expected, the longer the time series, the faster the algorithm converged.

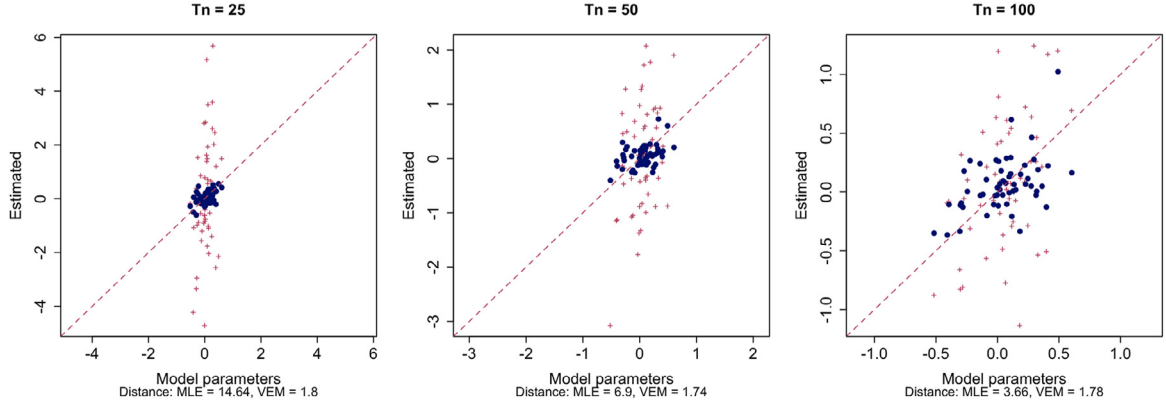

Fig. 3. An evaluation of MEM parameter estimation from the first subject ($n = 1$) among $N = 20$ virtually generated subjects with $d = 10$ ROI system. The result from model parameter estimation with MLE (red crosses) and VEM (navy dots) is displayed. The estimates from the two methods are compared with the true parameter that generated state sequences, where the x-axis is for the true parameters θ_1 and the y-axis is for the estimates $\hat{\theta}_1$. The Euclidean distance between the estimate and the model parameter for the subject $n = 1$ is also presented at the bottom of the scatterplot.

the pMEM parameters and the functional connectivity (FC, the Pearson correlation matrix of the independent components) for the three subnetworks.

To explore the relationship between the energy landscape features and the ADHD-related behavior scores, we conducted the sparse canonical correlation analysis (sCCA) (Chu et al., 2013), as was done in Kang et al. (2021)). For the ADHD-related behavioral scores, we used the subscales of the Child Behavior Checklist (CBCL) (CBCL, Achenbach and Rescorla, 2001) and the strengths and weaknesses of ADHD symptoms and normal behaviors (SWAN, Swanson et al., 2001), that are typically used to diagnose ADHD. More specifically, the CBCL score of attention problems (CBCL_AP), the CBCL score of internalizing problems (CBCL_INT), the CBCL score of externalizing problems (CBCL_EXT), the sum of inattentive item scores in SWAN (SWAN_INA), and the sum of hyperactive/impulsive item scores in SWAN (SWAN_HYP) were compared. Also, the age and the sex were included as the behavioral features.

Three different biological features were associated with the behavior scores using sCCA: (1) a vector of energy landscape features; (2) pMEM parameters (an upper triangle of $\boldsymbol{J}$); and (3) a vector of an upper triangle of functional connectivity. The energy landscape features are composed of features at the individual-level LMs, such as (1) the number of local minima (nLM), (2) OCC1, 3) OCC1+2, (3) the relative energy at the LM1 compared to the mean energy at all LMs (eLM1), (4) the relative average energy at the LM1 and LM2 minus mean energy at all LMs (eLM12), (5) the maximum energy minus mean energy at all LMs (eLMx1), (6) the maximum energy barrier from the LM1; and at the group level LMs, such as (7) the energy at the *i*-th group LM (eLMg*i*), (8) the average energies at the states within the distance (network size/5) from the *i*-th group LM (aeLMg*i*). LMg indicates the local minima of the group average energy landscape estimated from the group pMEM of all individuals. For the DMN and ECN+SAN, the number of LMg was 2 while 4 for ADHDN. For example, the DMN features include eLMg1, eLMg2, aeLMg1, and aeLMg2. To compare pMEM with FC in terms of the explanatory power for behaviors, we did not include the self-excitability in the analysis. All behavioral scores and neurobiological scores were normalized across subjects by the z-transformation.

3. Results

3.1. Simulation studies

Fig. 2 is the plot of ELBO evolving by each variational EM (abbr. VEM) step for the simulated data sets ($N = 20$, $d = 10$), which shows that the algorithm has converged well in all cases ($T_n = 25,\ 50,\ 100$). Furthermore, it is observed that, as expected, the longer the time series, the sooner the algorithm converged. Fig. 3 presents the scatterplots of the estimates versus the true parameters for the first subject ($n = 1$) among $N = 20$ virtually generated subjects from the simulation study, where the navy dots are for the VEM estimates and the red crosses are for the MLE. For all cases with $T_1 = 25,\ 50,\ 100$, it is clearly seen that the VEM gave better estimates than the MLE, and a much smaller distance of VEM estimates, printed at the bottom of each plot, also supports this observation. Most of all, the VEM yielded reasonable estimates regardless of the length T_1 of the time series, while the MLE seems to be collapsed when the time series is short. This was commonly observed all over the virtual subjects, although they have been omitted for saving space.

3.2. Monte Carlo simulation

Fig. 4 summarizes the result from our Monte Carlo simulation study for checking whether the finding from the above simulation studies can be generalized. For each simulation setup, the overall discrepancy ($\overline{D^2}$) was computed repeatedly 100 times during the iterative Monte Carlo experiments for both the MLE and the VEM. Each point in the figure depicts the pair of overall discrepancies of the MLE and the VEM corresponding to each Monte Carlo iteration. All the 100 points are laid in the

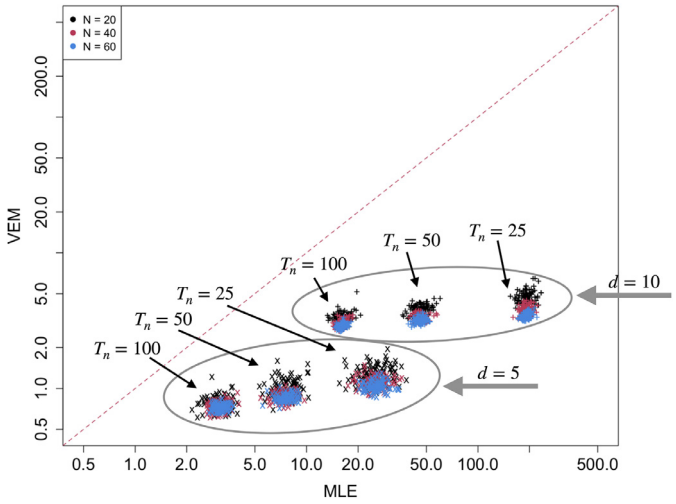

Fig. 4. The result from the Monte Carlo experiment. The overall discrepancy of the MLE and the VEM are compared for each iteration. The VEM yielded a more minor discrepancy than the MLE, although the discrepancy gap between them gets narrower as the length of the time series increases. The disparity of the MLE increased as the number of nodes increased while the performance of VEM had been affected little by the number of nodes.

right-bottom side of the straight line $y = x$, which shows that the proposed VEM yielded a more minor discrepancy than the MLE, although the discrepancy gap, as expected, gets narrower as the length of time series increases. Also, the disparity of the MLE got worse as the number of nodes increases while that of VEM has been affected little by the number of nodes.

Fig. 5 depicts the distribution of the ratios of overall discrepancies ($= \overline{D^2}_{\text{MLE}}/\overline{D^2}_{\text{VEM}}$). As can be inferred from Fig. 4, the ratios are much larger than 1 in all simulation settings, which indicates the VEM has defeated the MLE in all cases. And, the ratio tends to get more prominent as the number of subjects increases.

Additionally, we compared the proposed VEM procedure with another Bayesian procedure, BMEM, by Kang et al. (2021). Fig. 6 summarizes the result. Each point (+) in the first panel is the pair of overall discrepancy ($\overline{D^2}$) observed from 100 virtually generated datasets. All the 100 points are laid on the right-bottom side of the straight line $y = x$, which shows the VEM performed better than the BMEM consistently for all virtual datasets. The last two panels illustrate the discrepancies $D(\hat{\theta}_n,\ \theta_n)$ for each subject $n = 1, 2, \ldots, 20$ from the first two datasets among the 100 virtual datasets. Each bullet stands for the pair $(D_{\text{BMEM}}(\hat{\theta}_n,\ \theta_n),\ D_{\text{VEM}}(\hat{\theta}_n,\ \theta_n))$ of the discrepancies of the BMEM and the VEM for each subject, where we observed that the VEM gave better results for most subjects.

3.3. Experimental study

Fig. 7 shows independent components (IC) evaluated in the current study. The group differences of occupation times evaluated by ANCOVA are summarized in Table 1.

The ANCOVA revealed the significant group difference of the OCC1 at the DMN ($F = 6.996$, $p = 0.009$). The OCC1 of DMN in the ADHD (0.519 ± 0.007, mean $\pm$ s.d.) was lower than that of TDC (0.555 ± 0.012). The ANCOVA analysis shows the significant group differences of the OCC1+2 at the DMN ($F = 4.003$, $p = 0.019$) – shorter in ADHD (0.897 ± 0.009) than TDC (0.938 ± 0.015). For the ECN+SAN, ANCOVA of the OCC1+2 showed group differences ($F = 5.474$, $p = 0.020$) – longer in ADHD (0.886 ± 0.011) compared to TDC (0.837 ± 0.018). No significant group differences were found for the number of local minima at any subnetworks.

Table 1
Group differences of the energy landscape features and pMEM parameters analyzed by ANCOVA with age and sex as covariates.

Network	Features	ADHD Mean (std)	TDCMean (std)	Group effect F (p-value)
DMN	OCC1	0.519 (0.007)	0.555 (0.012)	6.996 (0.009)
	OCC1+2	0.897 (0.009)	0.938 (0.015)	5.619 (0.018)
ECN+SAN	OCC1+2	0.886 (0.011)	0.837 (0.018)	5.474 (0.020)
	H1 (IC6)	-0.642 (0.044)	-0.469 (0.075)	3.934 (0.048)
	J(IC7, IC8)	-0.164 (0.056)	0.069 (0.096)	4.444 (0.036)
ADHDN	J(IC4, IC11)	0.059 (0.055)	-0.178 (0.095)	4.676 (0.031)

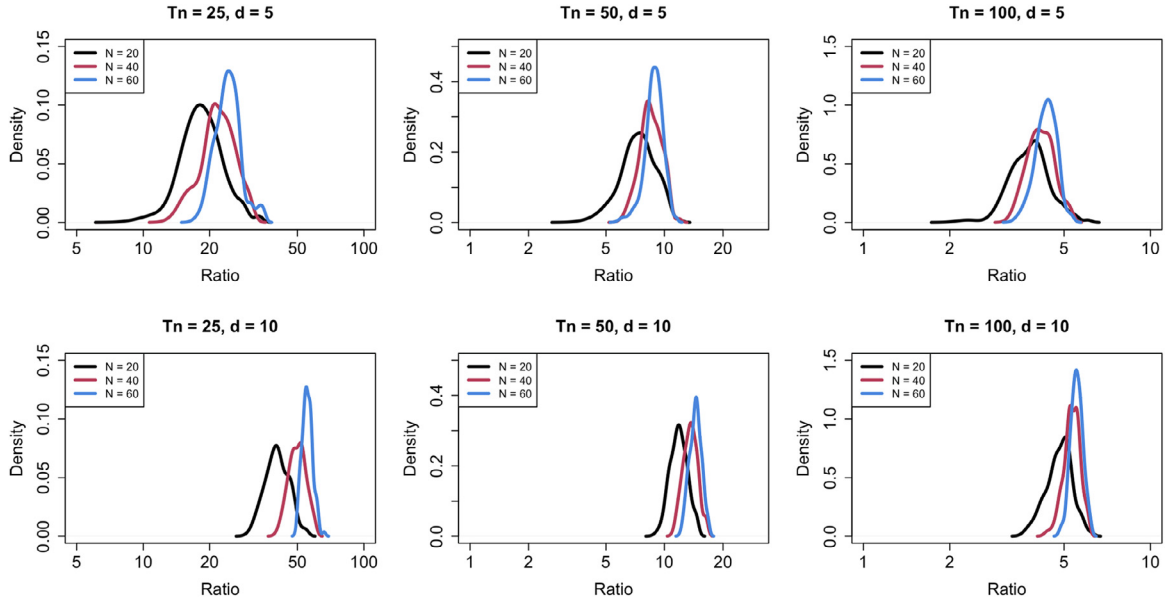

Fig. 5. The distributions of the ratio of overall discrepancies ($= \overline{D^2}_{\mathrm{MLE}} / \overline{D^2}_{\mathrm{VEM}}$). In all simulation settings, the ratios are much larger than 1, which indicates VEM has defeated MLE. And it is observed that the ratio tends to get larger as the number of subjects increases.

The ECN+SAN showed significantly lower excitability of IC6 in ADHD (-0.642 ± 0.044) compared to TDC (-0.469 ± 0.075) ($F = 3.934$, $p = 0.048$). The interaction between IC7 and IC8 at the ECN+SAN was anti-directional in the ADHD (-0.164 ± 0.056) in contrast to TDC (0.069 ± 0.096), the group difference of which was significant ($F = 4.444$, $p = 0.036$). The ADHDN showed group differences in the interaction between a component (IC4) in the DMN and a component in the SAN (IC11) ($F = 4.676$, $p = 0.031$), showing positive interaction in the ADHD (0.059 ± 0.055) compared to negative interaction in the TDC (-0.178 ± 0.095). There was no significant group difference in the functional connectivity at any subnetworks.

The results of sCCA for the energy landscape features, the pMEM parameters, and the FC are summarized in Table 2, and Fig. 8 shows two examples of sCCA results.

4. Discussion

The energy landscape analysis of the brain states has provided an important perspective in understanding the nature of the resting-state brain as a dynamic complex system (Ezaki et al., 2018; Gu et al., 2018; Kang et al., 2017; Watanabe et al., 2013, 2014b, 2014c, 2014d). For example, the resting-state brain is configured to have the maximal number of stable states when compared to other network configurations (Kang et al., 2017). It has a hub-like state transition organization (Kang et al., 2019). The energy landscape analyses have also demonstrated the role of energy landscape features in characterizing individuals (Kang et al., 2021) and groups such as autism spectrum disorders (Watanabe and Rees, 2017) and schizophrenia (Cabral et al., 2013; Loh et al., 2007). However, with a small dataset, pMEM-based energy landscape analysis would struggle to estimate its model parameters reliably.

To resolve this problem, Kang et al. (2021) previously proposed a Bayesian parameter estimation of the individual pMEM using the variational Bayes approximation method (BMEM). Kang et al. (2021) demonstrated that utilizing the group data information in estimating the individual model parameters as a prior improves the accuracy of estimation of the individual parameters. The BMEM assumes a zero-mean normal prior, although the prior mean vector is slightly modified by the concatenated group data by an iterative step. In the BMEM, the prior precision should be manually controlled; it uses an arbitrarily small positive number for the prior precision so that the resulting posterior mean vector does not shrink to the prior mean vector too much. They also presented another version of the BMEM that assumes a conjugate prior for the model parameters (see Appendix C of Kang et al. (2021)). However, the posterior of the model parameter of an individual reflects the group data information excessively, which results in little contribution of a new individual data in constructing its posterior, particularly for large group size.

This study successfully overcomes the limitations of the BMEM method by adopting the variational EM approach under the empirical Bayes framework. Unlike the BMEM, the VEM-MEM provides hyperparameter estimates as well as model parameter estimates in a fully empirical manner. The posterior for each subject explains the individualized pMEM sufficiently, even with a small sample size. Additionally, the normal prior distribution determined by the hyperparameter estimates, η and α, provides probabilistic information about the group under analysis because the hyperparameters are obtained by summarizing the individual model parameter estimates for all the subjects in the group. Hence,

Table 2
sCCA results for the ADHD, the TDC and the combined group of the ADHD+TDC at the DMN, the ECN+SAN and the ADHDN. Three different types of biological features are evaluated: EL indicates energy landscape features; and pMEM and FC indicate the upper triangles of the pMEM matrix(**J**) and the correlation matrix R. The seven (most important) elements of the first pair of canonical variables are displayed (sorted according to the weight strength of each element). ***r*** is the correlation coefficient of the first pair of canonical variables. See text for variable names.

Group	Network	Features	sCCAfirst pair of canonical variables	r
ADHD $N = 231$	DMN	EL	0.05 CBCL_AP + 0.05 SWAN_INA + 0.04 sex + 0.04 SWAN_HYP + 0.03 CBCL_INT + 0.02 age + 0.01 CBCL_EXT ∝ 0.30 aeLMg1 + 0.30 eLM12 + 0.19 aeLMg2 + 0.18 eLMg1 + 0.12 eLM1 + 0.12 nLM + 0.11 mxEB	0.363
		pMEM	0.04 age + 0.03 SWAN_HYP+ 0.03 CBCL_EXT+ 0.03 sex + 0.02 CBCL_INT+ 0.01 CBCL_AP + 0.01 SWAN_INA ∝ 0.04 J(1,2) + 0.03 J(2,3) + 0.03 J(4,5) + 0.02 J(1,5) + 0.02 J(3,4) + 0.02 J(1,3) + 0.02 J(2,5)	0.332
		FC	0.05 SWAN_HYP + 0.04 age + 0.03 CBCL_INT + 0.02 CBCL_EXT + 0.02 sex + 0.01 SWAN_INA + 0.00 CBCL_AP ∝ 0.07 R(1,2) + 0.05 R(2,3) + 0.04 R(4,5) + 0.03 R(1,3) + 0.03 R(3,4) + 0.03 R(2,5) + 0.03 R(1,4)	0.302
	ECN +SAN	EL	0.05 CBCL_INT + 0.04 sex + 0.03 SWAN_INA + 0.03 CBCL_EXT + 0.01 age + 0.00 CBCL_AP + 0.00 SWAN_HYP ∝ 0.20 eLM12 + 0.15 eLM1 + 0.09 eLMg2 + 0.08 aeLMg3 + 0.05 OCC1 + 0.04 eLMg3 + 0.04 eLMg1	0.333
		pMEM	0.04 sex + 0.04 SWAN_INA + 0.04 CBCL_AP + 0.03 CBCL_INT+ 0.03 SWAN_HYP+ 0.01 CBCL_EXT+ 0.00 age ∝ 0.03 J(2,5) + 0.03 J(4,6) + 0.02 J(1,4) + 0.02 J(1,2) + 0.02 J(1,5) + 0.02 J(2,6) + 0.02 J(1,6)	0.373
		FC	0.04 SWAN_INA + 0.04 CBCL_AP + 0.04 CBCL_EXT + 0.04 SWAN_HYP + 0.03 sex + 0.02 CBCL_INT + 0.01 age ∝ 0.04 R(2,4) + 0.04 R(1,6) + 0.03 R(2,5) + 0.03 R(1,4) + 0.03 R(4,6) + 0.02 R(1,2) + 0.02 R(3,4)	0.365
	ADHDN	EL	0.06 SWAN_INA + 0.05 CBCL_AP + 0.04 CBCL_EXT + 0.03 SWAN_HYP + 0.03 CBCL_INT + 0.02 age + 0.02 sex ∝ 2.14 aeLMg4 + 1.84 aeLMg1 + 1.68 eLMg4 + 1.46 eLMg1 + 0.52 aeLMg2 + 0.44 eLMg2 + 0.38 eLMg3	0.379
		pMEM	0.05 CBCL_INT+ 0.04 sex + 0.03 SWAN_HYP+ 0.03 CBCL_AP + 0.01 SWAN_INA + 0.00 CBCL_EXT+ 0.00 age ∝ 0.03 J(7,10) + 0.03 J(1,7) + 0.02 J(2,5) + 0.02 J(10,11) + 0.02 J(5,10) + 0.02 J(1,3) + 0.02 J(3,10)	0.621
		FC	0.04 CBCL_INT + 0.04 sex + 0.04 SWAN_HYP + 0.02 CBCL_AP + 0.01 SWAN_INA + 0.00 age + 0.00 CBCL_EXT ∝ 0.06 R(1,2) + 0.05 R(2,7) + 0.05 R(7,9) + 0.04 R(7,10) + 0.04 R(1,4) + 0.03 R(2,3) + 0.03 R(4,10)	0.604
TDC $N = 61$	DMN	EL	0.12 CBCL_AP + 0.09 CBCL_EXT + 0.06 age + 0.04 CBCL_INT + 0.04 sex + 0.03 SWAN_HYP + 0.01 SWAN_INA ∝ 0.69 aeLMg1 + 0.54 aeLMg2 + 0.34 eLMg2 + 0.31 eLMg1 + 0.28 eLM1 + 0.19 mxEB + 0.11 eLM12	0.7
		pMEM	0.10 CBCL_AP + 0.08 age + 0.08 CBCL_INT+ 0.02 sex + 0.01 SWAN_INA + 0.01 SWAN_HYP+ 0.01 CBCL_EXT ∝ 0.09 J(2,4) + 0.08 J(2,5) + 0.06 J(1,4) + 0.05 J(3,4) + 0.05 J(1,3) + 0.04 J(1,5) + 0.01 J(2,3)	0.577
		FC	0.14 CBCL_AP + 0.06 age + 0.05 SWAN_INA + 0.05 CBCL_INT + 0.01 sex + 0.01 CBCL_EXT + 0.00 SWAN_HYP ∝ 0.11 R(2,4) + 0.11 R(1,3) + 0.08 R(2,5) + 0.07 R(4,5) + 0.04 R(2,3) + 0.04 R(3,5) + 0.02 R(1,5)	0.651
	ECN +SAN	EL	0.10 SWAN_INA + 0.08 age + 0.08 SWAN_HYP + 0.07 sex + 0.06 CBCL_INT + 0.06 CBCL_EXT + 0.05 CBCL_AP ∝ 0.84 eLM12 + 0.52 eLM1 + 0.33 eLMg1 + 0.30 aeLMg4 + 0.22 aeLMg2 + 0.17 aeLMg1 + 0.16 eLMg2	0.624
		pMEM	0.12 age + 0.09 SWAN_HYP+ 0.07 SWAN_INA + 0.02 CBCL_EXT+ 0.01 CBCL_INT+ 0.00 CBCL_AP + 0.00 sex ∝ 0.07 J(2,3) + 0.06 J(3,5) + 0.05 J(3,4) + 0.05 J(2,5) + 0.05 J(5,6) + 0.05 J(1,5) + 0.03 J(2,4)	0.664
		FC	0.10 SWAN_HYP + 0.09 age + 0.06 CBCL_EXT + 0.04 CBCL_AP + 0.03 SWAN_INA + 0.02 sex + 0.01 CBCL_INT ∝ 0.11 R(3,4) + 0.09 R(2,4) + 0.08 R(1,3) + 0.08 R(1,4) + 0.08 R(3,5) + 0.08 R(1,5) + 0.06 R(5,6)	0.612
	ADHDN	EL	0.10 CBCL_INT + 0.10 CBCL_AP + 0.03 SWAN_INA + 0.03 age + 0.02 CBCL_EXT + 0.02 sex + 0.02 SWAN_HYP ∝ 0.68 aeLMg1 + 0.52 eLMg1 + 0.46 aeLMg4 + 0.33 eLMg4 + 0.23 eLMg2 + 0.22 OCC12 + 0.21 OCC1	0.672
		pMEM	0.13 SWAN_HYP+ 0.05 sex + 0.04 CBCL_INT+ 0.03 CBCL_AP + 0.03 age + 0.01 SWAN_INA + 0.01 CBCL_EXT ∝ 0.16 J(2,11) + 0.15 J(5,8) + 0.15 J(5,11) + 0.15 J(4,8) + 0.13 J(6,8) + 0.13 J(7,9) + 0.13 J(1,8)	1.000*
		FC	0.08 SWAN_HYP + 0.08 CBCL_EXT + 0.07 sex + 0.06 SWAN_INA + 0.05 age + 0.05 CBCL_INT + 0.02 CBCL_AP ∝ 0.30 R(4,7) + 0.23 R(1,5) + 0.22 R(3,5) + 0.20 R(1,6) + 0.19 R(7,8) + 0.19 R(3,9) + 0.17 R(7,10)	1.000*
ADHD TDC $N = 292$	DMN	EL	0.05 SWAN_INA + 0.05 SWAN_HYP + 0.04 CBCL_AP + 0.03 sex + 0.03 CBCL_INT + 0.02 CBCL_EXT + 0.01 age ∝ 0.35 aeLMg1 + 0.23 aeLMg2 + 0.19 eLMg1 + 0.13 eLMg2 + 0.12 eLM12 + 0.09 nLM + 0.07 OCC12	0.330
		pMEM	0.05 SWAN_HYP+ 0.04 CBCL_EXT+ 0.04 age + 0.03 SWAN_INA + 0.01 sex + 0.00 CBCL_INT+ 0.00 CBCL_AP ∝ 0.04 J(3,4) + 0.03 J(1,5) + 0.02 J(4,5) + 0.02 J(1,4) + 0.02 J(1,3) + 0.02 J(2,5) + 0.01 J(2,3)	0.276
		FC	0.05 SWAN_HYP + 0.05 CBCL_EXT + 0.03 age + 0.03 SWAN_INA + 0.02 sex + 0.00 CBCL_AP + 0.00 CBCL_INT ∝ 0.03 R(4,5) + 0.03 R(1,5) + 0.02 R(3,4) + 0.02 R(1,2) + 0.02 R(2,3) + 0.02 R(1,3) + 0.02 R(2,5)	0.277
	ECN +SAN	EL	0.04 SWAN_INA + 0.02 CBCL_EXT + 0.02 CBCL_INT + 0.01 SWAN_HYP + 0.01 sex + 0.00 CBCL_AP + 0.00 age ∝ 0.07 eLM1 + 0.07 eLM12 + 0.06 OCC12 + 0.04 eLMg4 + 0.03 aeLMg4 + 0.03 eLMg2 + 0.02 eLMg1	0.328
		pMEM	0.05 CBCL_AP + 0.04 CBCL_INT+ 0.04 sex + 0.03 SWAN_INA + 0.02 SWAN_HYP+ 0.01 age + 0.01 CBCL_EXT ∝ 0.04 J(4,6) + 0.03 J(1,4) + 0.02 J(1,5) + 0.02 J(2,5) + 0.02 J(1,2) + 0.02 J(2,3) + 0.02 J(2,4)	0.323
		FC	0.06 CBCL_EXT + 0.05 SWAN_INA + 0.05 CBCL_AP + 0.04 SWAN_HYP + 0.01 sex + 0.01 age + 0.00 CBCL_INT ∝ 0.04 R(2,4) + 0.03 R(1,4) + 0.03 R(3,4) + 0.03 R(4,6) + 0.02 R(1,6) + 0.01 R(2,5) + 0.01 R(5,6)	0.309

(*continued on next page*)

Table 2 (*continued*)

Group	Network	Features	sCCAfirst pair of canonical variables	r
	ADHDN	EL	0.06 SWAN_HYP + 0.03 sex + 0.03 CBCL_EXT + 0.02 SWAN_INA + 0.01 age + 0.01 CBCL_INT + 0.01 CBCL_AP ∝ 0.74 aeLMg4 + 0.70 aeLMg1 + 0.62 eLMg4 + 0.56 eLMg1 + 0.35 eLMg3 + 0.34 eLMg2 + 0.30 aeLMg3	0.340
		pMEM	0.05 CBCL_INT+ 0.03 sex + 0.03 SWAN_HYP+ 0.03 CBCL_AP + 0.01 age + 0.00 SWAN_INA + 0.00 CBCL_EXT ∝ 0.02 J(2,7) + 0.02 J(10,11) + 0.02 J(1,7) + 0.02 J(9,11) + 0.02 J(4,10) + 0.02 J(7,10) + 0.02 J(3,10)	0.527
		FC	0.03 CBCL_EXT + 0.03 sex + 0.03 CBCL_INT + 0.03 CBCL_AP + 0.02 age + 0.02 SWAN_HYP + 0.00 SWAN_INA ∝ 0.04 R(7,9) + 0.03 R(2,3) + 0.03 R(1,2) + 0.03 R(9,11) + 0.03 R(2,7) + 0.03 R(10,11) + 0.03 R(6,8)	0.495

the VEM-MEM enables every individual posterior to utilize the information from the other subjects in the group to ensure that it adequately reflects the essential concept of the empirical Bayes approach. The simulation studies with various setups support the arguments discussed so far and verify that the proposed variational EM algorithm converges adequately within a reasonable number of iterations.

Kang et al. (2021) presented the importance of the pMEM and the energy landscape features in characterizing individuals. The subject specificity of the pMEM parameters and their energy landscape features is highly comparable to that of the functional connectivity of the same dataset. They also reported that the local minima information of the individuals could be lost when a group model is constructed by the conventional method which uses the concatenated data across the individuals. In this study, the proposed VEM-MEM is applied to test the capability of characterizing the group differences between ADHD and TDC. The occupation time at the LM1 (OCC1) and the occupation time at both the LM1 and the LM2 (OCC1+2) in the DMN of ADHD are shorter than those of TDC. Meanwhile, the occupation time of the LM1 and LM2 at the ECN+SAN of the ADHD is longer than that of the TDC. The differences in the excitability at the component of the ECN (IC6) and the interaction between the components in the ECN (IC7 and IC8) when evaluated within the ECN+SAN were also found. Thus, the state dynamics in both the DMN and the ECN+SAN may characterize ADHD. Shappell et al. (2021) reported that children with ADHD spent less in the states where the DMN is anticorrelated with other task-related regions (e.g., ECN+SAN) but spent more in the states where the DMN is in positive synchrony with other task-related regions compared to the TDC. Suppose we associate the findings of Shappell et al. (2021) with the current result. In that case, LM1 at the DMN may be associated with the anticorrelated state between the DMN and other subnetworks, while LM1 and LM2 at the ECN+SAN may be associated with the highly correlated state between the DMN and other task-related regions. Since this study aimed to show the plausibility of the VEM-MEM in characterizing group differences, the ADHD characteristics in the nonlinear dynamics will be analyzed in detail in future studies.

The interaction between the components in the DMN and the SAN of the ADHDN (IC4 and IC11) showed the anti-directional group differences between the ADHD and the TDC. Nevertheless, no group differences of the energy landscape feature were found at the ADHDN. It confirms the nonlinear nature of the brain system – adding subnetworks generates a different network. No significant group difference of energy landscape features despite the pMEM difference can be explained by the nonlinear relationship between pMEM and energy landscape features. It is also possible that the more complex energy landscapes that emerged from a more extensive network could be more heterogeneous across the individuals. In line with the relationship between the pMEM and the energy landscape features, the pMEM parameters at the DMN, which indicate the interactions among the brain regions or the self-excitability, are not significantly different between the two groups. However, the energy landscape feature such as the occupation times differed. Although the energy landscape features are derived from the pMEM parameters, they directly reflect the information about the frequent brain states and their properties in the dynamics, which cannot be explored from the interaction parameters in the pMEM.

The pMEM shows the behavioral characteristics more sensitively than the FC features measured by the Pearson correlation matrix. The ANCOVA failed to find any significant group differences in the correlation-based conventional functional connectivity. Furthermore, in the sCCA, the correlation levels of the behavioral features with the pMEM parameters were generally higher than those with the functional connectivity. Although the pMEM and the functional connectivity cannot be compared directly since they explain the different aspects of the

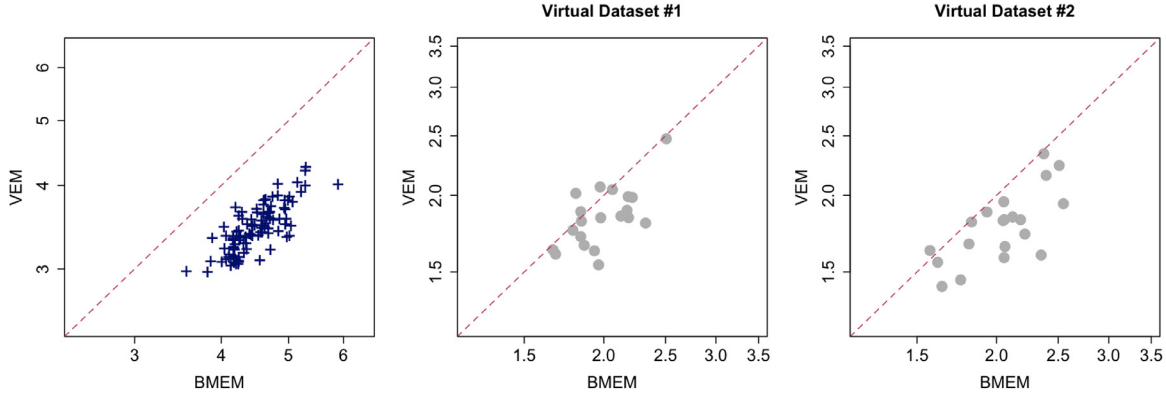

Fig. 6. Comparison of the estimation accuracy between VEM and BMEM. The first panel compares the overall discrepancies observed from 100 Monte Carlo simulations. Each point (+) is the pair of overall discrepancy $(D^2_{\text{BMEM}},\ D^2_{\text{VEM}})$. All the 100 points are laid in the right-bottom side of the straight line $y = x$, which shows VEM performed better than BMEM for all virtually generated datasets. The last two panels illustrate the discrepancies $D(\hat{\theta}_n,\ \theta_n)$ for each subject $n = 1, 2, \ldots, 20$, observed from the first two datasets among the 100 virtual datasets. Each bullet stands for the pair of discrepancies of BMEM and VEM for each subject, where we observed that the proposed VEM gave better parameter estimates for most subjects.

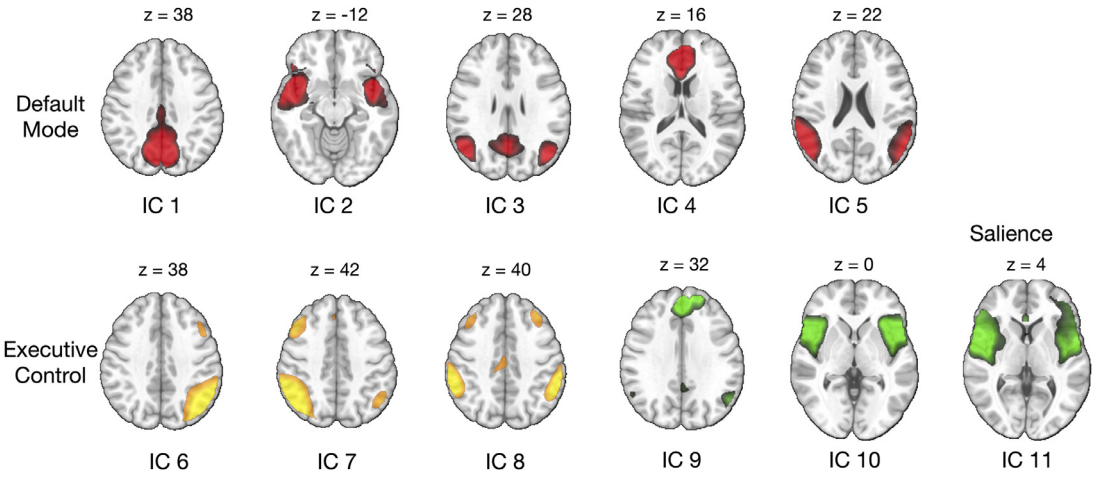

Fig. 7. Core ADHD-related components used in the current study.

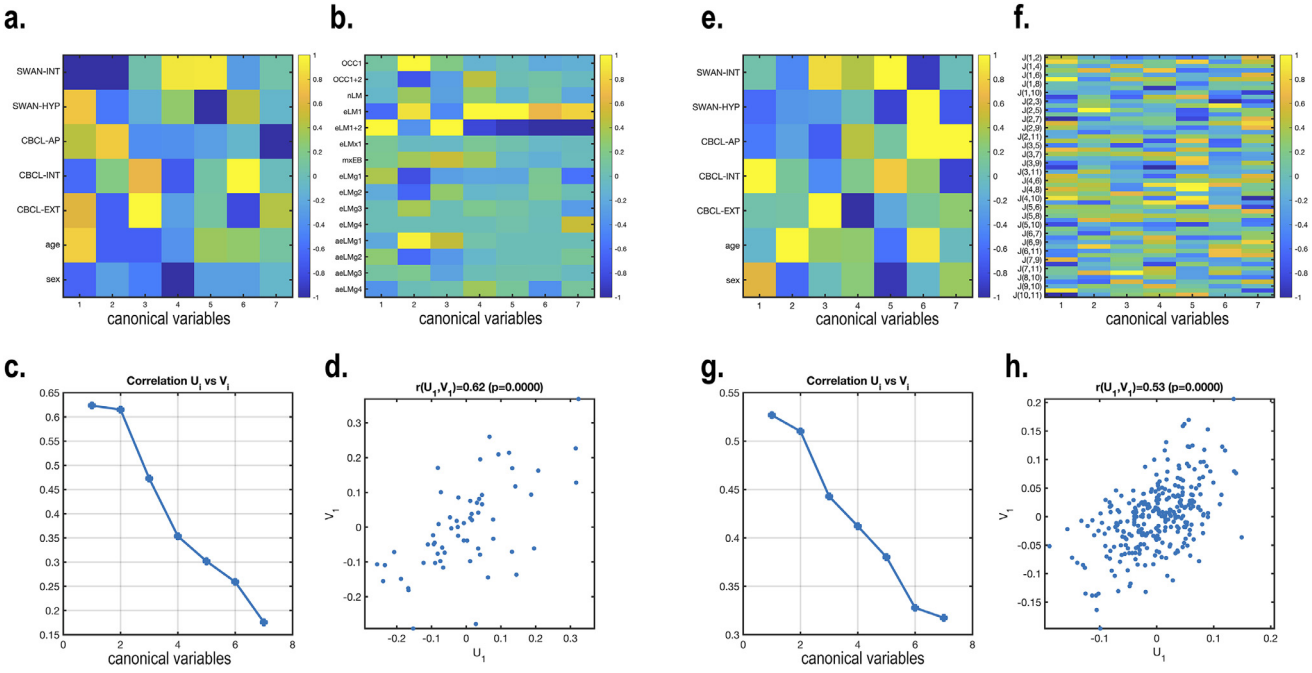

Fig. 8. Examples of sCCA between behavior scores and energy landscape features at the ADHDN in TDC (a-d) and pMEM parameters (J) at the ADHDN in ADHD+TDC (e–h). a) canonical variables for behavior scores (U, with gender and age), b) canonical variables for energy landscape feature (V), c) canonical correlations between every pairs of canonical variables for behavior scores (each column in a) and energy landscape features (the column in b, corresponding to the column in a) are displayed. d) The individual samples of behavior scores and canonical variables for energy landscape features projected onto the first canonical variable for cognition score (U_1) and the first canonical variable for energy landscape features (V_1) showed a correlation of $r = 0.62$. e) - h) are same as a) - d) except for using pMEM parameters (J) instead of energy landscape features.

interactions, this differential finding suggests the need for the individualized pMEM as much as the individualized FC. Both simulation and experimental studies support the plausibility of the VEM-MEM in characterizing individuals.

In the current study, we adopted the group-pooling approach. If the group were composed of sufficiently many human subjects, the prior derived from the group data would guide robust estimates by keeping the individual parameter estimates from taking erratic values. In this situation, the choice of the group prior may not matter. We speculate that the group-specific prior could be advantageous in estimating model parameters of an individual in a large group with relatively small inter-individual variations compared to inter-group variations. However, if the inter-individual variation is high within a group, parameter estimation with group-specific prior may not be significantly different from the prior derived by pooling groups. When some groups in the dataset are relatively small, the group-pooling approach would be a better choice. Furthermore, when we need to compare parameters statistically across groups, we do not want a bias in the individual parameter estimate vector due to the group distinction. It happens very often that the group labeling has uncertainty, and the biological grouping is required. In this case, the group-pooling approach could be practical again.

In the present study, we tested pMEM estimation with a network size of around ten, which belongs to network size ranges of widely used subnetworks. In theory, the current approach could be extended to estimate pMEM for a larger network with tens of nodes by adopting a pseudo-likelihood (Ezaki et al., 2017). Practical issues with pMEM for a larger network are the reliability of the model parameter estimation and impracticality to derive energy landscape properties due to the ex-

ponentially growing state dimension. The need to further study on the individualized large-scale energy landscape analysis from our Bayesian framework for pMEM estimation of a larger network remains.

As a concluding remark, we point out that the current study is somewhat technical and that further research is required using basic and clinical settings that can validate the usefulness of the proposed method and expedite our understanding of the brain in terms of nonlinear dynamics.

Credit authorship contribution statement

Seok-Oh Jeong: Conceptualization, Methodology, Software, Writing – original draft. **Jiyoung Kang:** Software, Methodology, Formal analysis. **Chongwon Pae:** Data curation, Formal analysis. **Jinseok Eo:** Data curation, Formal analysis. **Sung Min Park:** Data curation, Validation. **Junho Son:** Data curation. **Hae-Jeong Park:** Conceptualization, Methodology, Writing – original draft, Project administration, Funding acquisition.

Acknowledgment

This research was supported by Brain Research Program and the Korea Research Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C7A1030750 and NRF-2017M3C7A1049051). Seok-Oh Jeong's work was partially supported by the research fund of Hankuk University of Foreign Studies (2021).

Data code

Data used in the current study was downloaded from the open database. The codes in this study are available from (https://github.com/monetyonsei/paper).

References

Achenbach, T.M., Rescorla, L.A., 2001. Manual for the ASEBA School-age Forms & Profiles: An Integrated System of Mult-informant Assessment. ASEBA.

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Dufek, S., Eaves, M., Fradera, B., Gardner, J., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Horton, M., Kahn, D., Kabotyanski, K., Karmel, B., Kelly, S.P., Kleinman, K., Koo, B., Kramer, E., Lennon, E., Lord, C., Mantello, G., Margolis, A., Merikangas, K.R., Milham, J., Minniti, G., Neuhaus, R., Levine, A., Osman, Y., Parra, L.C., Pugh, K.R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F.X., Klein, A., Paus, T., Leventhal, B.L., Craddock, R.C., Koplewicz, H.S., Milham, M.P., 2017. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data 4, 170181.

Anderson, A., Douglas, P.K., Kerr, W.T., Haynes, V.S., Yuille, A.L., Xie, J., Wu, Y.N., Brown, J.A., Cohen, M.S., 2014. Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. Neuroimage 102 (Pt 1), 207–219.

Barber, A.D., Jacobson, L.A., Wexler, J.L., Nebel, M.B., Caffo, B.S., Pekar, J.J., Mostofsky, S.H., 2015. Connectivity supporting attention in children with attention deficit hyperactivity disorder. Neuroimage Clin. 7, 68–81.

Breakspear, M., 2017. Dynamic models of large-scale brain activity. Nat. Neurosci. 20, 340–352.

Brown, M.R., Sidhu, G.S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P.H., Greenshaw, A.J., Dursun, S.M., 2012. ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. Front. Syst. Neurosci. 6, 69.

Cabral, J., Fernandes, H.M., Van Hartevelt, T.J., James, A.C., Kringelbach, M.L., Deco, G., 2013. Structural connectivity in schizophrenia and its impact on the dynamics of spontaneous functional networks. Chaos Interdiscip. J. Nonlinear Sci. 23, 046111.

Cabral, J., Kringelbach, M.L., Deco, G., 2014. Exploring the network dynamics underlying brain activity during rest. Prog. Neurobiol. 114, 102–131.

Cai, W., Chen, T., Szegletes, L., Supekar, K., Menon, V., 2015. Aberrant cross-brain network interaction in children with attention-deficit/hyperactivity disorder and its relation to attention deficits: a multisite and cross-site replication study. Biol. Psychiatry.

Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14, 140–151.

Cao, Q., Zang, Y., Sun, L., Sui, M., Long, X., Zou, Q., Wang, Y., 2006. Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. Neuroreport 17, 1033–1036.

Chu, D., Liao, L.Z., Ng, M.K., Zhang, X., 2013. Sparse canonical correlation analysis: new formulation and algorithm. IEEE Trans. Pattern Anal. Mach. Intell. 35, 3050–3065.

Deco, G., Jirsa, V.K., 2012. Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. J. Neurosci. 32, 3366–3375.

Deco, G., Tononi, G., Boly, M., Kringelbach, M.L., 2015. Rethinking segregation and integration: contributions of whole-brain modelling. Nat. Rev. Neurosci. 16, 430–439.

Ezaki, T., Fonseca Dos Reis, E., Watanabe, T., Sakaki, M., Masuda, N., 2020. Closer to critical resting-state neural dynamics in individuals with higher fluid intelligence. Commun. Biol. 3, 52.

Ezaki, T., Sakaki, M., Watanabe, T., Masuda, N., 2018. Age-related changes in the ease of dynamical transitions in human brain activity. Hum. Brain Mapp. 39, 2673–2688.

Ezaki, T., Watanabe, T., Ohzeki, M., Masuda, N., 2017. Energy landscape analysis of neuroimaging data. Philos. Trans. A Math Phys. Eng. Sci. 375.

Fair, D.A., Posner, J., Nagel, B.J., Bathula, D., Dias, T.G., Mills, K.L., Blythe, M.S., Giwa, A., Schmitt, C.F., Nigg, J.T., 2010. Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. Biol. Psychiatry 68, 1084–1091.

Freyer, F., Roberts, J.A., Becker, R., Robinson, P.A., Ritter, P., Breakspear, M., 2011. Biophysical mechanisms of multistability in resting-state cortical rhythms. J. Neurosci. 31, 6353–6361.

Freyer, F., Roberts, J.A., Ritter, P., Breakspear, M., 2012. A canonical model of multistability and scale-invariance in biological systems. PLoS Comput. Biol. 8, e1002634.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 20, 189–210.

Gu, S., Cieslak, M., Baird, B., Muldoon, S.F., Grafton, S.T., Pasqualetti, F., Bassett, D.S., 2018. The energy landscape of neurophysiological activity implicit in brain network structure. Sci. Rep. 8, 2507.

Hilger, K., Fiebach, C.J., 2019. ADHD symptoms are associated with the modular structure of intrinsic brain networks in a representative sample of healthy adults. Netw. Neurosci. 3, 567–588.

Kang, J., Jeong, S.O., Pae, C., Park, H.J., 2021. Bayesian estimation of maximum entropy model for individualized energy landscape analysis of brain state dynamics. Hum. Brain Mapp..

Kang, J., Pae, C., Park, H.J., 2017. Energy landscape analysis of the subcortical brain network unravels system properties beneath resting state dynamics. Neuroimage 149, 153–164.

Kang, J., Pae, C., Park, H.J., 2019. Graph-theoretical analysis for energy landscape reveals the organization of state transitions in the resting-state human cerebral cortex. PLoS One 14, e0222161.

Kelso, J.A., 2012. Multistability and metastability: understanding dynamic coordination in the brain. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 906–918.

Konrad, K., Neufang, S., Hanisch, C., Fink, G.R., Herpertz-Dahlmann, B., 2006. Dysfunctional attentional networks in children with attention deficit/hyperactivity disorder: evidence from an event-related functional magnetic resonance imaging study. Biol. Psychiatry 59, 643–651.

Li, Y.-O., Adalı, T., Calhoun, V.D., 2007. Estimating the number of independent components for functional magnetic resonance imaging data. Hum. Brain Mapp. 28, 1251–1266.

Loh, M., Rolls, E.T., Deco, G., 2007. A dynamical systems hypothesis of schizophrenia. PLoS Comput. Biol. 3, e228.

Park, H.J., Kang, J., 2021. A computational framework for controlling the self-restorative brain based on the free energy and degeneracy principles. Front. Comput. Neurosci..

Qiu, M.G., Ye, Z., Li, Q.Y., Liu, G.J., Xie, B., Wang, J., 2011. Changes of brain structure and function in ADHD children. Brain Topogr. 24, 243–252.

Rabinovich, M.I., Varona, P., 2011. Robust transient dynamics and brain functions. Front. Comput. Neurosci. 5, 24.

Saad, J.F., Griffiths, K.R., Korgaonkar, M.S., 2020. A systematic review of imaging studies in the combined and inattentive subtypes of attention deficit hyperactivity disorder. Front. Integr. Neurosci. 14, 31.

Scofield, J.E., Johnson, J.D., Wood, P.K., Geary, D.C., 2019. Latent resting-state network dynamics in boys and girls with attention-deficit/hyperactivity disorder. PLoS One 14, e0218891.

Shappell, H.M., Duffy, K.A., Rosch, K.S., Pekar, J.J., Mostofsky, S.H., Lindquist, M.A., Cohen, J.R., 2021. Children with attention-deficit/hyperactivity disorder spend more time in hyperconnected network states and less time in segregated network states as revealed by dynamic connectivity analysis. NeuroImage 229, 117753.

Sidlauskaite, J., Sonuga-Barke, E., Roeyers, H., Wiersema, J.R., 2016. Altered intrinsic organisation of brain networks implicated in attentional processes in adult attention-deficit/hyperactivity disorder: a resting-state study of attention, default mode and salience network connectivity. Eur. Arch. Psychiatry Clin. Neurosci. 266, 349–357.

Sonuga-Barke, E.J., Castellanos, F.X., 2007. Spontaneous attentional fluctuations in impaired states and pathological conditions: a neurobiological hypothesis. Neurosci. Biobehav. Rev. 31, 977–986.

Swanson, J., Deutsch, C., Cantwell, D., Posner, M., Kennedy, J., Barr, C., Moyzis, R., Schuck, S., Flodman, P., Spence, A., 2001. Genes and attention-deficit hyperactivity disorder. Clin. Neurosci. Res. 1, 207–216.

Tognoli, E., Kelso, J.A., 2014. The metastable brain. Neuron 81, 35–48.

Watanabe, T., Hirose, S., Wada, H., Imai, Y., Machida, T., Shirouzu, I., Konishi, S., Miyashita, Y., Masuda, N., 2013. A pairwise maximum entropy model accurately describes resting-state human brain networks. Nat. Commun. 4, 1370.

Watanabe, T., Hirose, S., Wada, H., Imai, Y., Machida, T., Shirouzu, I., Konishi, S., Miyashita, Y., Masuda, N., 2014a. Energy landscapes of resting-state brain networks. Front. Neuroinf. 8, 12.

Watanabe, T., Hirose, S., Wada, H., Imai, Y., Machida, T., Shirouzu, I., Konishi, S., Miyashita, Y., Masuda, N., 2014b. Energy landscapes of resting-state brain networks. Front. Neuroinform. 8, 12.

Watanabe, T., Kan, S., Koike, T., Misaki, M., Konishi, S., Miyauchi, S., Miyahsita, Y., Masuda, N., 2014c. Network-dependent modulation of brain activity during sleep. Neuroimage 98, 1–10.
Watanabe, T., Masuda, N., Megumi, F., Kanai, R., Rees, G., 2014d. Energy landscape and dynamics of brain activity during human bistable perception. Nat. Commun. 5, 4765.
Watanabe, T., Rees, G., 2017. Brain network dynamics in high-functioning individuals with autism. Nat. Commun. 8, 16048.
Yeh, F.C., Tang, A.N., Hobbs, J.P., Hottowy, P., Dabrowski, W., Sher, A., Litke, A., Beggs, J.M., 2010. Maximum entropy approaches to living neural networks. Entropy 12, 89–106.
Zang, Y.F., He, Y., Zhu, C.Z., Cao, Q.J., Sui, M.Q., Liang, M., Tian, L.X., Jiang, T.Z., Wang, Y.F., 2007. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 29, 83–91.